Comparison of Various Machine Learning Regression Models Based on Human Age Prediction

Manaf K. H. AlTaleb[1],

Affiliations
[1]Department of Electrical Engineering, Wasit University, Wasit, Iraq

Correspondence
Manaf K. H. AlTaleb,
Department of Electrical Engineering, Wasit University, Wasit, Iraq
Email: Manaf@uowasit.edu.iq

Abstract

The development of machine learning strategies has made it possible to diagnose some disease automatically based on data obtained from medical imaging. Brain age is one of the factors that can be used as an indicator of cognitive well-being. Recent advancements in machine learning have made it possible for computers to anticipate classification and prediction outcomes more accurately than humans.

In this study, five widely used machine learning regression models (Linear support vector regression (L-SVR), radial basis function support vector regression (RBF-SVR), relevance vector regression (RVR), Elastic Net and Gaussian process regression (GPR)) were trained and evaluated to predict brain age using volumes of brain regions data. Moreover, a dimensionality reduction technique was utilized to reduce the dimensionality of the input feature space. The data were collected from one hundred and eleven participants.

The results showed no performance difference amongst models trained on the same type of data, suggesting that the type of input data had a stronger influence on prediction performance than the model choice. The experimental results indicated that the GPR was the best fit model (R2=0.57, R=0.75) among the other regression models while the G-SVR was the worst fit model (R^2=0.0006, R=0.025) with such number of the input data.

الخلاصة: ان التطور الحاصل في مجال التعلم الالي جعل من الممكن تشخيص بعض الأمراض تلقائيًا بناءً على البيانات التي تم الحصول عليها من التصوير الطبي. يعد عمر الدماغ أحد العوامل التي يمكن استخدامها كمؤشر على الرفاهية المعرفية. حيث من خلال معرفة عمر الدماغ ومقارنته مع العمر الحقيقي للفرد يمكن معرفة صحة عمرالدماغ وامكانية الاصابة بمصطلح ما يعرف بالشيخوخة المتقدمة او المتأخرة والتي تعتبر كمؤشر حيوي للكثير من الوظائف الادراكية. لقد جعلت التطورات الحديثة في التعلم الآلي من الممكن لأجهزة الكمبيوتر توقع النتائج بشكل أكثر دقة من البشر.في هذه الدراسة ، تم استخدام خمسة نماذج مستخدمة على نطاق واسع من تحليل الانحدارالموجودة ضمن انظمة التعلم الآلي وهي انحدار ناقل الدعم الخطي (L-SVR) ، استخدام خاصية وظيفة الأساس الشعاعي التي تدعم الانحدارالمتجه (RBF-SVR) ، انحدار متجه الصلة (RVR)، الشبكة المرنة وانحدارالعملية الغاوسية(GPR). تم تدريب هذه النماذج للحصول على تنبؤ لعمرالدماغ البشري باستخدام بيانات ماخوذة من مناطق الدماغ. علاوة على ذلك، تم استخدام تقنية تقليل الخصائص لتقليل خصائص بيانات الإدخال. تم جمع البيانات من مئة واحد و عشر مشاركا. أظهرت النتائج عدم وجود فرق في الأداء بين النماذج المدربة على نفس النوع من البيانات، مما يشير إلى أن نوع بيانات الإدخال كان له تأثير أقوى على الأداء من اختيار النموذج. تشير النتائج إلى أن GPR كان أفضل نموذج ملائم للتنبوء ($R^2 = 0.57$, $R = 0.75$)   من بين نماذج الانحدار الأخرى بينما كان G-SVR هو النموذج الأسوأ من حيث دقة التنبؤ ($R = 0.025$, $R^2 = 0.0006$) مع هذا العدد من المدخلات بيانات.

1. INTRODUCTION

Globally, aging and its accompanying health issues provide a significant burden to individuals and organizations. Cognitive decline and an increased risk of neurodegenerative illness are also factors linked with aging of the brain; however, the severity of these effects varies widely between individuals. There is a rapid increase in the average age of the world's population, which is expected to reach 17% globally by the year 2050 and 25% in Europe and North America by the same year [1]. The growing proportion of elderly people in the population is associated with higher overall social and economic expenses [1]. Therefore, it is of the utmost importance to avoid or slow down the progression of further deterioration of such age-related health problems in their earlier stages and reduce the indirect cost as well. Because of that, efforts are constantly growing to address the early stages of this challenge [2]. Brain age prediction is a method for measuring the effects of aging on the brain, based on the well-established link between age and neuroanatomy over the lifecycle. Brain age gap refers to the difference between an individual's anticipated age and their chronological age. It is the primary outcome metric in brain age prediction. A positive brain age gap, which is commonly denoted as accelerated or premature aging, indicates that a person's anticipated brain age is older than their real age. A negative brain age gap, which is sometimes referred to as delayed aging, suggests a younger expected brain age [3].To determine whether the terms accelerated or delayed aging are appropriate, however, more research into the neurobiological underpinnings of brain aging is necessary. Machine learning (ML) algorithms have made it possible to automatically predict diseases using medical imaging data [4,5]. In some situations, the current advancement of ML pushes prediction accuracy beyond human ability and can help with clinical diagnostic and treatment decisions [6,7]. In the field of neuroimaging, ML has proven successful in a number of applications for predictive and diagnostic analytics, including modelling and estimating the age of the brain [8-10]. Recently, there has been a lot of interest in assessing how the brain ages through the use of ML techniques for predicting brain age, which are often based on structural magnetic resonance imaging (MRI) [3,9]. Several researchers have found strong relationships between brain age forecasts and chronological age [8,11]. Jiang et al. (2020) introduced CNN-based age prediction models for seven structural brain networks images of healthy persons. Furthermore, they assessed how well CNN performed in terms of age estimation when compared to GPR and RVR ML algorithms. The result showed that the CNN is outperform the performance of GPR and RVR models in age prediction. They suggested that their CNN model can be possibly used for brain age prediction to diagnose disorders [12]. Peng H, et al. (2020) introduced a lightweight deep learning network for brain age prediction using T1-weighted MRI data. The algorithm was combined with other methods to improve the performance. They compared their approach with other broadly used machine learning techniques. The results revealed that the introduced model was successfully able to predict brain age and sex clarification [13]. Baecker L, et al. (2021) studied the performance of different ML approaches in voxel-based and regional data set. It suggested that the voxel-based models achieved the overall best performance using PCA; however, this strategy was computationally demanding and may not be practical if computational or time is restricted. Whereas a training set of only 120 participants was sufficient for region-based RVR to achieve satisfactory results, also this type was quick and easy to implement [14]. Baecker L, et al. (2021) introduced the concepts underlying brain age prediction using machine learning and discussed its potential clinical applications. It suggested that the use of such technologies may facilitate earlier and more precise treatment of age-related diseases. It may support differential diagnosis, prediction, and therapy decisions as well as early detection of brain-based illnesses [3]. Ganaie et al. (2022) assessed the brain age prediction using different ML regression techniques. The suggested frameworks involved twenty-two algorithms. The best accuracy was achieved using Quadratic SVR algorithm (R^2=0.88) and the lowest was achieved using Binary Decision Tree technique (R^2=0.76), suggesting that the accuracy of the prediction can be improved using advanced ML algorithms [15]. Ramírez et al. (2022) compared age prediction accuracy of extreme gradient boosting ML model and partial least squares ML model. The first algorithm achieved lower mean absolute error in predicting age than the second algorithm. The authors indicated that the cortical thickness in temporal parietal lobes exhibited better prediction accuracy than frontal and occipital lobes. Additionally, they suggested that integrating the prediction model and interpretation process could help to reduce the gap between chronological and real brain age [16].

Predictive models of brain age developed from brain imaging can determine the apparent biological age of a person's brain based on brain shape and/or function using ML techniques. It is possible for a person's brain age to be significantly different from their chronological age, and the age at which individual trajectories begin to deviate from the norms of the population can indicate important aspects of a person's brain health at any point in their life, from infancy to old life. Although many studies have been achieved in this field, there are still a lot of conceptual and technical obstacles to overcome in order to effectively anticipate a person's brain age and correctly interpret age gaps in the brain.

The selection of appropriate ML methods, particularly when working with small dataset, as well as appropriate neuroimaging features are among the primary challenges. In this study, five widely-used ML techniques (GPR, Elastic Net, L-SVR, RBF-LVR, RVR) for estimating brain age were studied. Moreover, an implementation of a feature selection was examined to determine whether it will help to improve the prediction accuracy for each approach. Therefore, the effect of the mentioned ML approaches on structural MRI data-trained models were evaluated.

2. METHODOLOGY

2.1. Participant

The total number of the participants where 122. One hundred and eleven participants (63 M,48F) with a mean age±SD= 49±17.3 as shown in figure 1, with normal or corrected to normal vision participated in the experiment. Eleven participants with contraindications to MRI (e.g., metal implants), claustrophobia, pregnant, tumour, or neurological disorders were excluded.
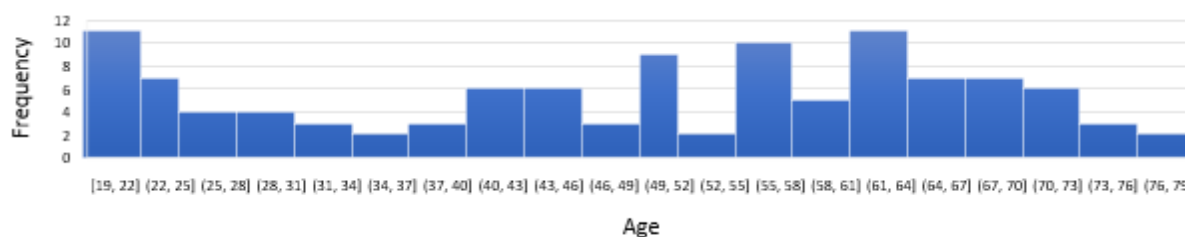

Figure 1 Age Distribution of 111 Participants.

2.2. Imaging Parameters

Participants were scanned in a 3 Tesla Siemens MAGNETOM MRI scanner at the neurosurgery teaching hospital-Baghdad. T1 weighted structural scans were acquired with the following parameters (TR=4.6ms, TE1.4ms, 192 sagittal slices, 1 mm^3 isotropic voxels and image resolution 256×256) as shown in figure 2.

Figure 2 Sagittal, Axial, and Coronal Views of Anatomical Image of One Participant.

2.3. Data Pre-processing

All the structural raw data were preprocessed using Freesurfer software package, v5.0 (http://surfer.nmr. mgh.harvard.edu/). The technical details of this software package are described in [17]. Briefly, the preprocessing involves:

- Motion correction: Correcting for subject head motion inside the MRI scanner.
- Intensity normalization: Correct the WM intensity.
- Talairach registration: Normalize the subject's brain to a standard MNI template.
- Skull stripping: Remove the non-brain structure, e.g. skull, soft tissues, and CSF.
- Cortical and Subcortical segmentation and labelling: Segmenting and identifying the cortical and subcortical regions according to Destrieux Atlas.
- Segmentation of white matter (WM): Labelling the WM structure.
- Tessellation of the GM/WM and GM/CSF boundaries: Identify the GM/WM, and GM/SCF.
- Cortical surface reconstruction: Reconstructed the WM surface, and inflate it as shown in figure 3.

Each participant's cortical surface was parceled into 74 gyral-based regions of interest (ROIs) per hemisphere according to the Destrieux Atlas as shown in figure 4 [18]. Cortical and subcortical (8 ROIs per hemisphere) GM volumes (in mm^3) were calculated at the ROI level of each participant.
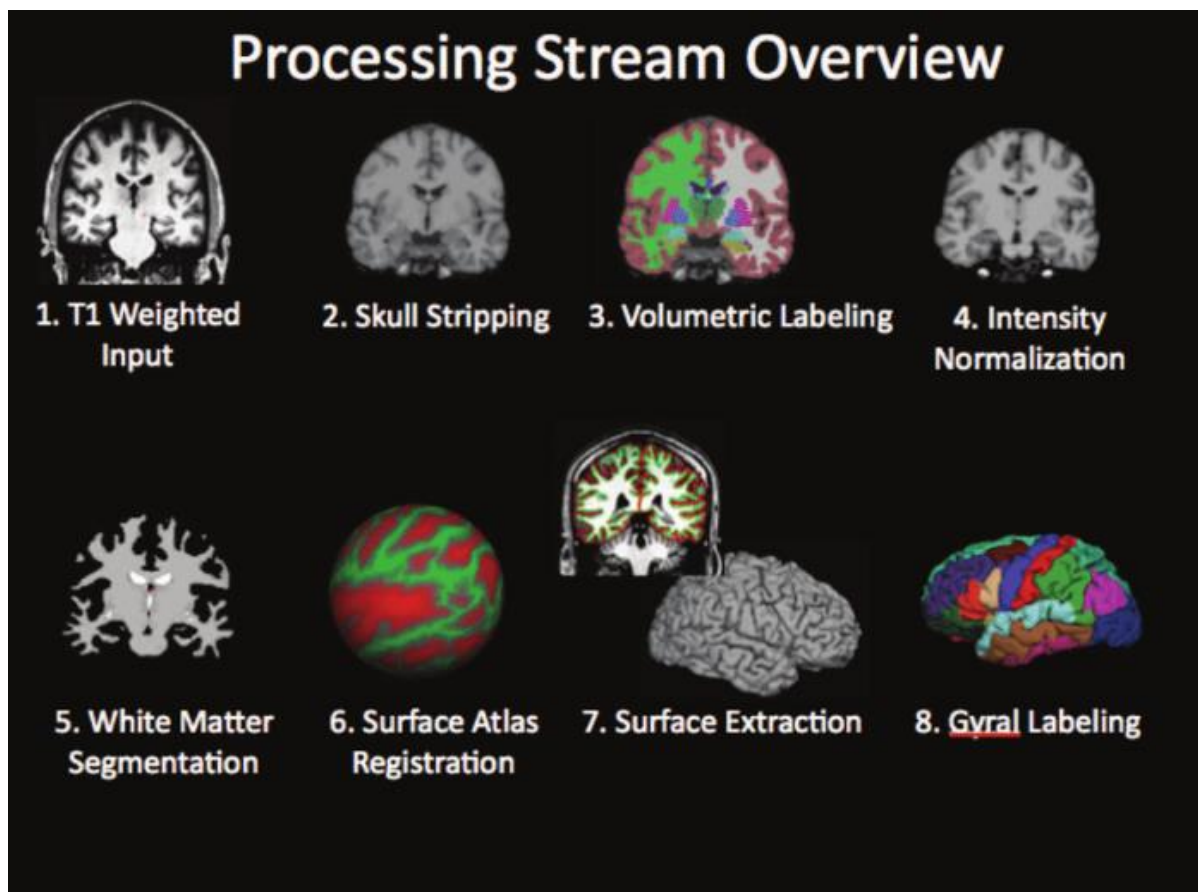

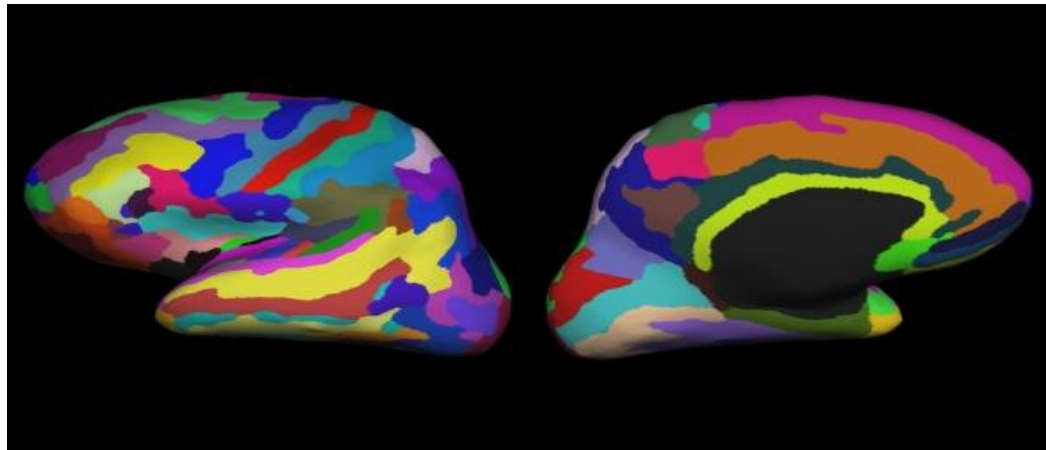**Figure** 3 Freesurfer pre-processing pipeline [17].

Figure 4 Lateral and medial views of Destrieux Atlas [18].

2.4. Data Analysis

All fMRI data were further analysed using MATLAB (R2019a), MathWorks, LibSVM package, and RVR package.

2.5. Data Preparation

After pre-processing all the structural MRI images, 162 ROI volumes were extracted from each participant, and divided by the total intracranial volume (TIV) to account for different brain sizes and to normalize the volume data across all participants. Participants' (one-dimensional) volumetric data were stacked, forming a two-dimensional feature matrix of (111subj X 162 features).

2.6. Prediction Algorithm

The prediction algorithm consists of two stages, feature selection using a filter algorithm (Regression ReliefF) followed by a regression algorithm (L-SVR, RBF-SVR, RVR, GPR, or Elastic Net) as shown in figure 5.
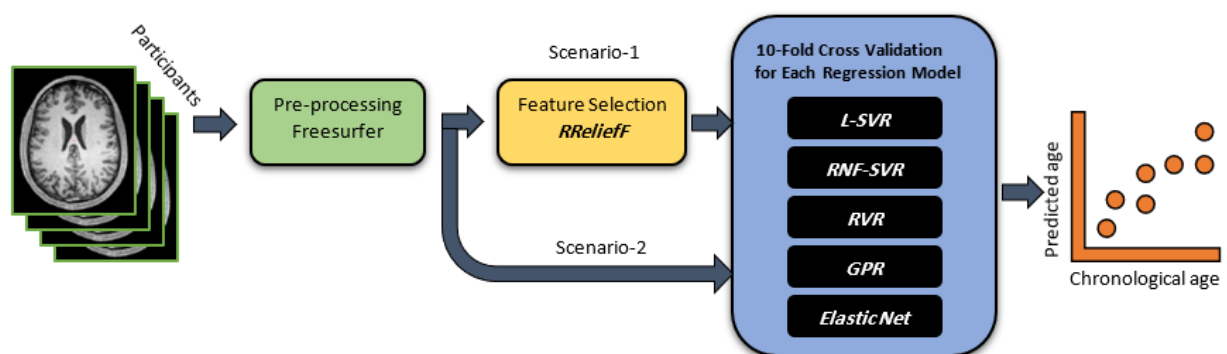

Figure 5 Schematic Diagram of the Prediction Framework.

2.6.1. Regression ReliefF

Relief algorithms are successful, generalized attribute estimators, which have been utilized in a many setting [19]. They can detect conditional connections between characteristics and give a unified view of the estimation of attributes in regression and classification [20]. It has traditionally been seen as feature subset selection methods that are used prior to the model being developed to estimate the quality of characteristics in regression models [21]. It penalizes predictors who assign different values to neighbours with identical response values and rewards predictors who assign different values to neighbours with different response values. Intermediate weights are used to compute the final predictor weights for predictor F_j.

$$W_j = \frac{Wdy \wedge dj}{Wdy} - \frac{Wdj - Wdy \wedge dj}{m - Wdy} \qquad (1)$$

where:

W_j = Predictor wights.

m = No. of iterations

W_{dy} = Weight of different values for y.

W_{dj} = Weight of different values for F$_j$.

$W_{dy \wedge dj}$ = Weight of different response and different values for the predictor F$_j$.

The RReliefF function is available in MATLAB help for direct implementation and more explanation.

2.6.2. Linear Support Vector Regression (L-SVR)

The support vector machine (SVM) is a supervised learning technique that uses labelled training date set to build input-output mapping functions [22]. A classification or a regression function can be used as the mapping function. SVMs have shown extremely competitive performance in a variety of real-world applications, including bioinformatics, neuroscience, face recognition, and image processing, establishing SVMs as one of the cutting-edge methods for machine learning and data mining, alongside other soft computing techniques. SVMs can be characterized as linear or nonlinear [14, 23]. The linear SVR model seeks a flat hyperplane that deviates as little as possible from the training data. It differs from linear regression in that the model aims to reduce observed training errors. The SVR calculates error only on data items that are outside of a margin of tolerance [24]. The hyperparameter epsilon (ε) determines the margin of tolerance, which establishes the maximum deviation from the hyperplane that data set can have [23]. The data points that located outside of this margin are known as support vector (SV) because they control the location of the hyperplane. The regularisation hyperparameter (C) is another factor that determines SVR performance. This parameter is used to decrease overfitting by balancing the hyperplane complexity with the obtained training errors [23,24]. The formula that can be used for prediction new data using SV:

$$f(x) = \sum_{n=1}^{N} (a_n - a_n^*)(x_n' x_n) + b \qquad (2)$$

with condition of $0 \leq a_n \leq C$

$$0 \leq a_n^* \leq C$$

where:

N= No. of training points

a_n, a_n^* = Lagrange multipliers (nonnegative real numbers).

x_n = Training points.

b = Training points.

C = Regularisation hyperparameter.

2.6.3. Nonlinear Support Vector Regression (RBF-SVR)

Sometimes, linear model cannot appropriately address some regression situations. In this scenario, the linear SVM technique can be extended to nonlinear functions using the Lagrange formulation [24]. A nonlinear SVM regression model can be obtained by swapping the dot product $x_1' x_2$ with a nonlinear kernel function φ (x_1,x_2) that maps x to a high-dimensional space [24].

$$f(x) = \sum_{n=1}^{N}(a_n - a_n^*)\,\varphi\big(xx_n\big) + b \tag{3}$$

where:

N = No. of training data

a_n, a_n^* = Lagrange multipliers (nonnegative real numbers).

x$_n$= Training points.

b = Bias.

φ = Kernel funnctio (Gaussian).

2.6.4. Relevance Vector Regression (RVR)

A Relevance Vector Machine (RVM) is a machine learning technology that employs Bayesian inference to provide efficient solutions for different regression and classification problems. The RVM has the same functional form as the SVM, nonetheless it can perform probabilistic regression. It is essentially a process of using Gaussian approach with a specific covariance model. In comparison to SVM, the Bayesian function of the RVM avoids the SVM's set of free parameters.

RVR employs a broad linear model of Bayesian formulation, which results in probabilistic rather than deterministic predictions [25]. The latter is accomplished by assuming a zero-mean normal distribution for the weights of the input data and iteratively modifying the precision values through approximation process. Weights with low precision are set to zero during training, and the basic functions related with them are trimmed. RVR results are often sparser than SVR results, implying that they employ fewer SV, which donates to their superior robustness to outliers and higher generalisation. Additionally, because RVR does not require hyperparameter adjustment, it eliminates the requirement for approaches such as random or grid search, potentially making the RVR training procedure simpler with less computational time [26]. Though, because the learning approach is a variant of expectation maximisation, the optimisation is non-convex, making predictions more susceptible to local minima mistakes [26].

$$k(x, x') = \sum_{j=1}^{N}\frac{1}{\alpha_j}\varphi\big(x, x_j\big)\varphi\big(x', x_j\big) \tag{4}$$

where:

N = No. of training vectors

α_j = Variance.

x…x$_j$= are the input vectors.

φ = Kernel funnctio (Gaussian).

2.6.5. Gaussian Process Regression (GPR)

GPR is a supervised machine learning model that represents a nonparametric Bayesian inference to a standard regression [27]. It is commonly used for modelling of indefinite functions or surfaces in wide range of regression and classification to spatial processes [28]. GPR deduces a probability distribution of probable values rather than learning the specific target value of training data. GPR needs the specification of a prior distribution in the form of a mean and covariance. It is commonly believed to be a multivariate normal Gaussian distribution with a mean of zero. Using Bayes' theorem, the probabilities of this prior distribution are then changed based on the goal values in the training data. The data from the previous distribution and the actual data are blended into joint probabilities in the ensuing posterior distribution. The predictive distribution for previously unnoticed data will be standard Gaussian if the previous distribution is believed to be Gaussian. The forecast for a previously unknown value can be inferred as the mean from this prediction of the distribution, and the prediction's uncertainty as its variance, therefore the RVR is a sparse GPR with a predefined covariance [29].

$$k(x, x') = \sigma_f^2 \, exp[\frac{-(x - x')^2}{2l^2}] \qquad (5)$$

where:

l = Length between two points

σ_f = Covariance.

x = Input vector.

2.6.6. Elastic Net

Elastic Net is a regression method that simultaneously does variable selection and regularization. It's a kind of modified linear regression. Regularization is essential when the problem of overfitting occurs. Overfitting occurs when the dataset contains several features, some of which are irrelevant to the prediction model [30]. This makes the model more complicated and its test set prediction too wrong (or overfitting problem). A model with such a high variance cannot generalise to fresh data [31]. To address these concerns, both L2 and L1 norm regularization can be included to obtain the advantages of Ridge and Lasso (Least Absolute Shrinkage and Selection Operator) regression simultaneously. The resulting model is more accurate in making predictions than Lasso [30,32]. It does feature selection while also simplifying the hypothesis. Here is the adjusted cost function for Elastic-Net model [33]:

$$= \frac{1}{m}\left[\sum_{l=1}^{m}(y^i - h(x^i))^2 + \lambda_1 \sum_{j=1}^{n}(w_j) + \lambda_2 \sum_{j=1}^{n}(w_j^2)\right] \qquad (6)$$

where:

m = Total number of training samples.

h(x(i)) = Hypothetical function for prediction.

y(i) =The value of target variable for ith training example.

w(j) = Weight for jth feature.

n = No. of features in the dataset.

λ_1 = Regularization strength for L1 norm.

λ_2 = Regularization strength for L2 norm

2.7. Prediction Performance

There are numerous methods used to assess the performance of the regression model (prediction error) in predicting data such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) [34]. It has been stated that all distance measurements (MSE, RMSE and MAE) are equivalent and assistance to quantify the precision of the approximated result compared to the simulated data [31]. Moreover, most of the studies of healthy people typically assess the accuracy of a brain age prediction model in terms of mean absolute error (MAE) [3,13]. In this study, we adopted MAE and RMSE which are defined as

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (7)$$

$$RMSE = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (8)$$

where n is the number of samples, y_i is the actual value and $\hat{y}_i$ is the predicted value. This criterion measures the Mean Absolute Error driven by the test data. thus, a low MAE or RMSE value means that the predicted age values are close to the real values [34].

2.8. Prediction Framework

In this study, the prediction framework was performed using MATLAB. It utilized the MRI volumetric data either with a feature selection stage (scenario 1) or without (scenario 2) to predict the participant's chronological age. A train-test split (90%-10%) was performed with a 10 folds cross validation to evaluate the generalizability of the regression model using the testing split data set as shown in figure 6. In Scenario-1, the training dataset was used as an input to the RReliefF algorithm to identify the most relevant features in the dataset. Next, the positively weighted features shown in figure Y were selected as a training subset to train a regression model. Then, a regression model (L-SVR, RBF-SVR, RVR, GPR, or Elastic Net) was built using the training subset (scenario 1) or the training set (scenario 2) and evaluated using the testing subset (scenario 1) or the testing set (scenario 2). Subsequently, the MAE / RMSE (overall prediction performance) were calculated as an average of the 10 folds performance.
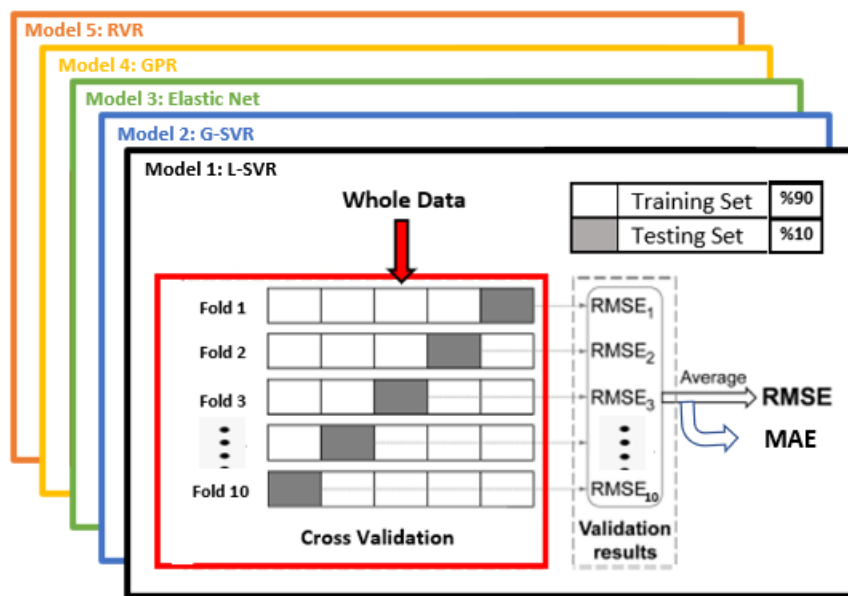

Figure 6 10 Folds Cross Validation with 90% Training Set from Whole Data and 10% Testing Set.
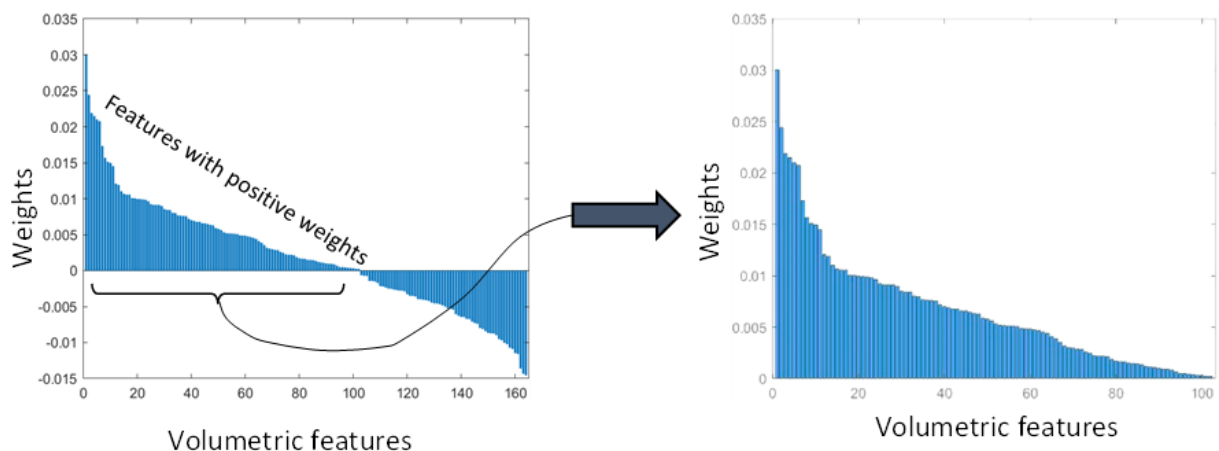

Figure 7 Output of RReliefF Algorithm Shows the Features Weights (left). The Positive Weighted Features were Selected to Train a Regression Model (right).

3. RESULTS

3.1 .Scenario 1: Prediction with feature selection

A different number of optimal features were selected for each of the 10th folds as illustrated in table 1. These features were used to train a regression model and test its performance. Five different regression models were tested and compared as shown in figure 8. It can be clearly seen that the Elastic Net and GPR models were the best in predicting the age compared to the others.
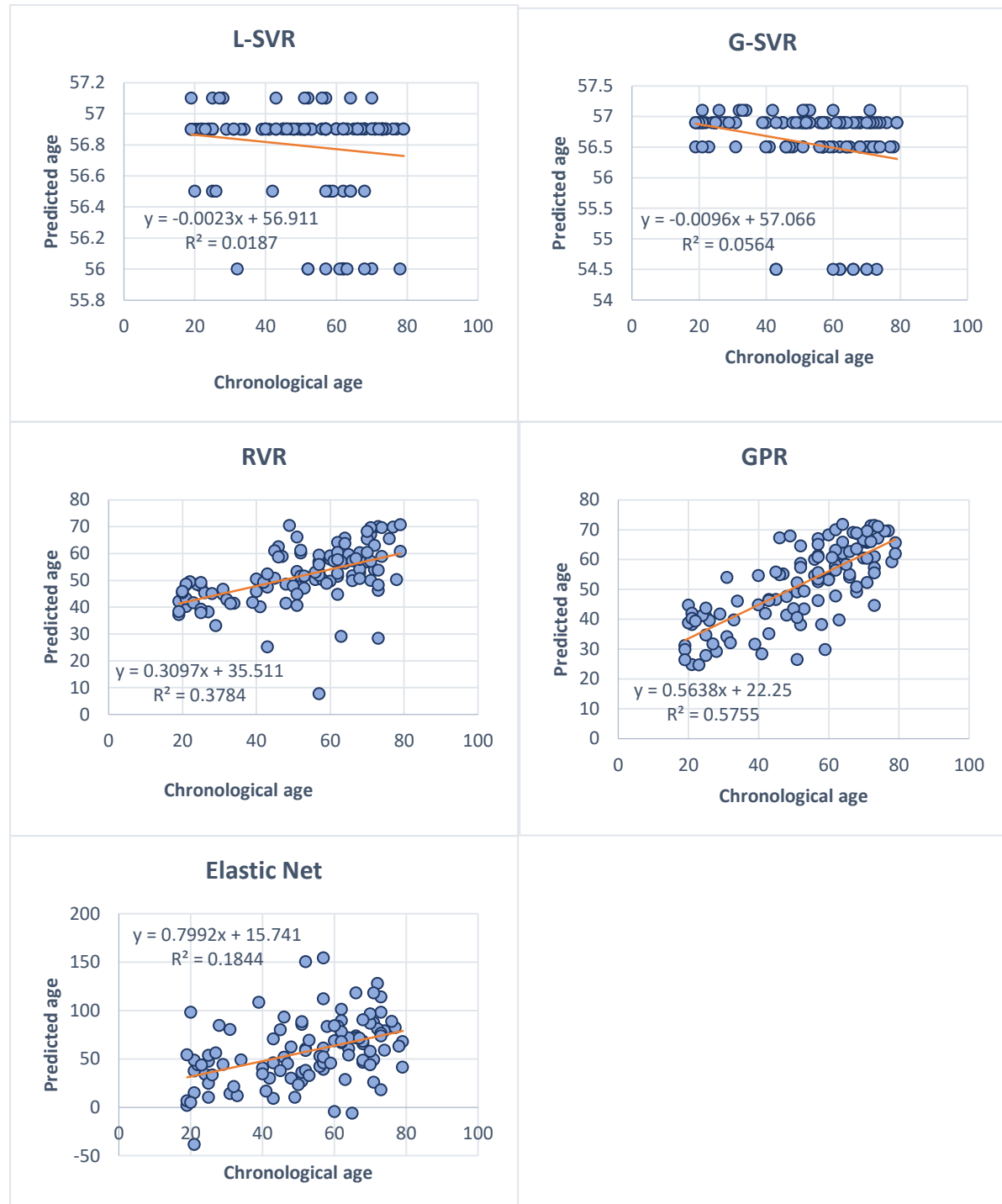

Figure 8 Scatter plots and correlation coffecients (R^2) show the chronological age (years) and the predicted age (years) using RReliefF as a feature selection approach and five different regression models. Highest accuracy was observed when training and testing using GPR followed by RVR models using a reduced set of features for the testing dataset

Table 1. Number of Positively Weighted Features Selected by RReliefF

Method	No. of Features in Each Fold									
	1	2	3	4	5	6	7	8	9	10
L-SVR	96	103	99	96	96	108	103	73	109	100
G-SVR	92	109	111	90	105	98	108	90	83	102
ELASTIC NET	99	97	100	103	96	104	102	96	83	101
GPR	90	91	92	97	97	102	92	105	104	103
RVR	100	108	88	102	101	102	99	97	83	102

3.2 .Scenario 2: Prediction without feature selection

In the second scenario, regression models were trained and tested using the entire feature set (162 volumetric features). Similarly, GPR and Elastic Net models achieved accurate age prediction compared to L-SVR, RBF-LVR and RVR as shown in figure 9.
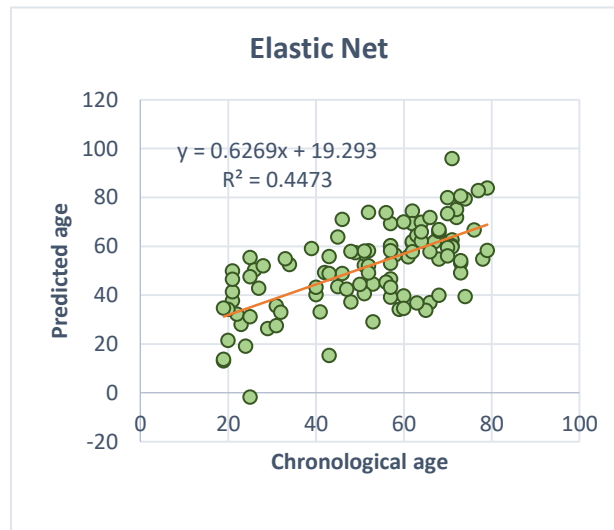
Figure 9 Scatter plots and correlation coffecients (R^2) show the chronological age (years) and the predicted age (years) using five different regression models. Highest accuracy was observed when training and testing using GPR followed by Elastic Net and RVR models using all the 162 volumetric features for the testing dataset.

3.3 Models performance comparison

Figure 10 depicts the age prediction accuracy MAE for these 5 regression models for the testing dataset with feature reduction approach and without. Further details regarding additional performance measures (R, R^2, RMSE) are shown in figures 11,12 and 13) respectively.
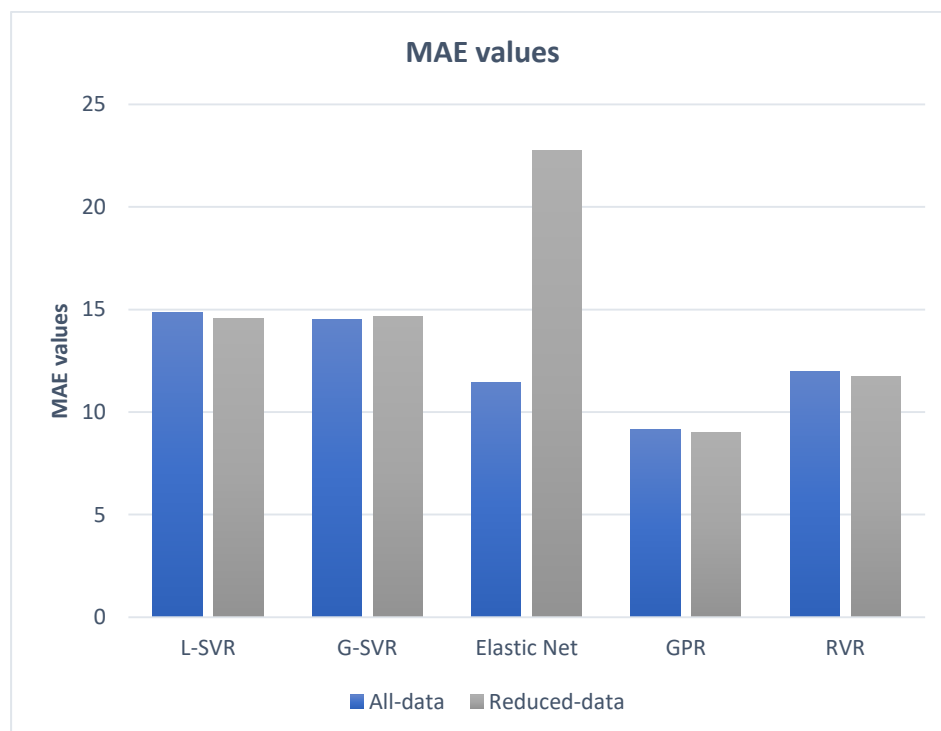

Figure 10 Age prediction MAE of L-SVR, RBF-SVR, Elastic Net, GPR, and RVR with Feature Selection (gray) and Without (blue) for the testing dataset.
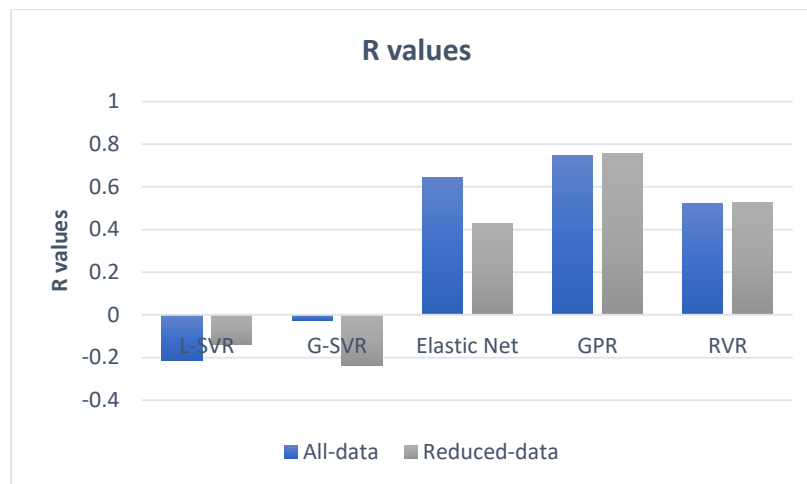
Figure 11 Age prediction R values of L-SVR, RBF-SVR, Elastic Net, GPR, and RVR with Feature Selection (gray) and Without (blue) for the testing dataset.
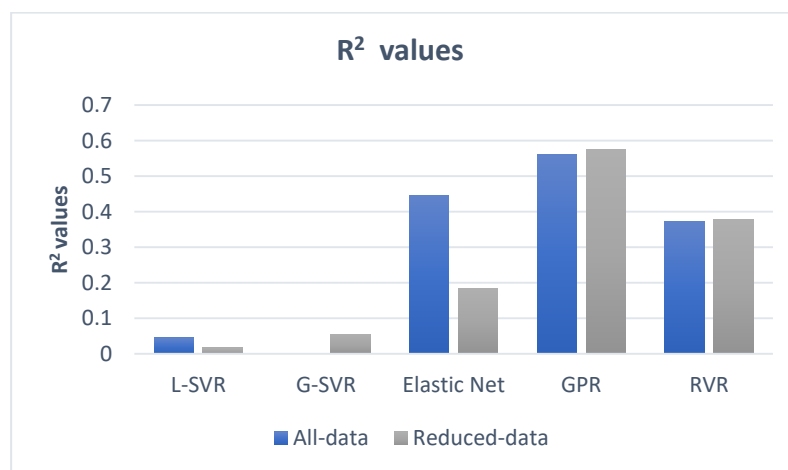

Figure 12 Age prediction R^2 values of L-SVR, RBF-SVR, Elastic Net, GPR, and RVR with Feature Selection (gray) and Without (blue) for the testing dataset.
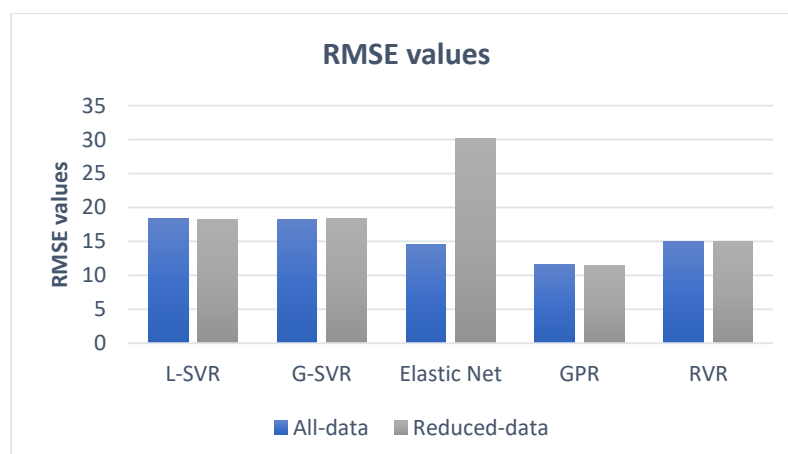

Figure 13 Age prediction RMSE values of L-SVR, RBF-SVR, Elastic Net, GPR, and RVR with Feature Selection (gray) and Without (blue) for the testing dataset.

4. DISCUSSIONS

The present study compared five regressions models L-SVR, RBF-SVR, Elastic Net, GPR, and RVR with different input features to predict brain age. The assessment performed in a total of 5 models separated into two scenarios.

The first scenario includes using L-SVR, RBF-SVR, Elastic Net, GPR, and RVR after utilizes of RReliefF to reduce the number of the suggested futures. Thus, through using only the futures that have the positive weighted as explained in figure 7. The higher accuracy of brain age prediction during this setup was achieved by GPR (R^2 =0.57, R=0.76) followed by RVR (R^2 =0.37, R=0.52). Moreover, the other models were unable to achieve acceptable prediction rate.

The second scenario includes using L-SVR, RBF-SVR, Elastic Net, GPR, and RVR with all 162 suggested futures. The higher accuracy of brain age prediction was achieved by again GPR (R^2 =0.56, R=0.75) followed by Elastic Net (R^2 =0.44, R=0.65) and RVR (R^2 =0.37, R=0.52). Moreover, the other two models were unable to reach adequate forecast rate.

The L-SVR and G-SVR regression model were unable to obtain acceptable performance (R^2 =0.04, R=-0.1), (R^2 =0.0006, R=-0.04) and (R^2 =0.018, R=-0.2), (R^2 =0.05, R=-0.24) respectively in both scenario in comparable with other implemented models, same observation was utilized by [13].

The GPR regression model was the best in the age predicting compared to the others in the both scenarios which is in line with result obtained by [13, 32] and [33]. The MAE and RMSE value were the same in both scenarios indicating that the reduction algorithm in the feature selecting was ineffective for GPR model. Meanwhile, the presence of feature selection algorithm had a negative impact on Elastic Net regression model, were the MAE and RMSE value drift from 11.4 to 22.7 and 14.5 to 30.2 respectively. This could be due to eliminating a number of some effective features [14]. Moreover, feature selection algorithm had no effect on the performance of GPR, L-SVR and G-SVR prediction-based models. However, the prediction accuracy has been improved using RVR regressing model implanting the RReliefF feature selecting algorithm. It seemed that dimensionality reduction algorithm could successfully remove some redundant characteristics in the RVR region-based model.

In conclusion, we assume that the difference between all-data and reduced for GPR and RVR is negligible, however for Elastic Net, using all-data produced higher performance due to the importance of including all features in building the regression model. Furthermore, reducing the features space (no. of features) may eliminate some of the relevant features that are part of the input feature pattern which is necessary for training the regression model.

The main limitation of this study was the limited number of the data set that was gathered form the hospital. However, the features of this particular data collection may affect our recommendations for sample size and processing capacity. Nevertheless, the achieved results were close to the result obtained by [14] with 120 samples training data. We do think that other data sets can benefit from applying the basic principle that certain models demand significantly more training data and processing resources than others.

While the present study assessed a variety of machine learning model and data input methodological options, there are a number of other methods that might be explored in the future. These models may include some age correction in the training data set. Several kinds of correction have been introduced recently [37-40]. In Addition, a better performance may be achieved through multi-modal data set such as combining region level and voxel-level data features rather than using single model futures. Thus, making them a possible path for future study on brain age prediction models.

5. CONCLUSIONS

Predicting a person's chronological age with acceptable precision using machine learning model trained on T1-MRI is possible in healthy people. This can be done using limited raw MRI data, with just a minimum amount of processing required to produce a reliable age estimate. The GPR algorithm was the best regression model among the other regression methods that have been presented in this study with boundary of the input data. In the future,

we may be able to use brain-predicted age as a biomarker to learn more about the aging brain and its connection to cognitive decline, neurodegeneration, and age-related brain illnesses using multimodal input data.

AKNOWLDGEMENTS

REFERENCES

1. United Nations, Department of Economic and Social Affairs, P. D. (2019). World population prospects 2019: Highlights (Issue ST/ESA/SER.A/**423**). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/ 12283219.

2. Efraim Jaul and Jeremy Barron. (2017). 'Age-Related Diseases and Clinical and Public Health Implications for the 85 Years Old and Over Population'. *Frontiers in Public Health*, Dec.2017. pubmed.ncbi.nlm.nih.gov/29312916/. doi: 10.3389/fpubh.2017.00335

3. Lea Baeckera, Rafael Garcia-Diasa, Sandra Vieiraa, Cristina Scarpazzaa,b, Andrea Mechellia, (2021). 'Machine learning for brain age prediction: Introduction to methods and clinical applications'. *EBioMedicine* **72** ,103600, https://doi.org/10.1016/j.ebiom.2021.103600

4. LeCun, Y., Bengio, Y., Hinton, G., (2015). Deep learning. *Nature* 521, 436 – 444. https: //doi.org/10.1038/nature14539

5. Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., Yang, Y., Guo, G., Xiao, M., Du, M., Qu, X. (2018). "Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment'. Alzheimer's Disease Neuroimaging Initiative, T.A.D.N., *Front. Neurosci*. **12**, 777,. https://doi.org/ 10.3389/fnins.00777.

6. Liu, M., Zhang, J., Adeli, E., Shen, D.,(2018). 'Landmark-based deep multi-instance learning for brain disease diagnosis'. *Med. Image Anal*. **43**, 157–168. https://doi. org/10.1016/j.media.2017.10.005 .

7. Kohl, S.A. A ., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J.R., Maier-Hein, K.H., Ali Eslami, S.M., Rezende, D.J., Ronneberger, O. (2018), 'A probabilistic U-net for segmentation of ambiguous images.,*Advances in Neural Information Processing Systems*. pp. 6965–6975.

8. Cole, J. H., & Franke, K. (2017). 'Predicting age using neuroimaging: Innovative brain ageing biomarkers'. *Trends in Neurosciences*, 40(**12**),681–690. https://doi.org/10.1016/j.tins.2017.10.001

9. Kawahara, J., Brown, C.J., Miller, S.P., Booth, B.G., Chau, V., Grunau, R.E., Zwicker, J.G., Hamarneh, G., (2017). 'BrainNetCNN: convolutional neural networks for brain networks; towards predicting neurodevelopment'. *Neuroimage* **146**, 1038–1049. https://doi.org/10.1016/j.

10. Cole, J. H., Franke, K., & Cherbuin, N. (2019). Quantification of the biological age of the brain using neuroimaging. 'Biomarkers of human aging. *Healthy ageing and longevity*'. Vol. 10, pp. 293–328. Cham, Switzerland: Springer.

11. Smith, S.M., Elliott, L.T., Alfaro-Almagro, F., McCarthy, P., Nichols, T.E., Douaud, G., Miller, K.L., (2020). 'Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations'. *Elife* 9, 802686. https://doi.org/10.7554/eLife.52677.

12. Jiang H, Lu N, Chen K, Yao L, Li K, Zhang J and Guo X, (2020)." Predicting Brain Age of Healthy Adults Based on Structural MRI Parcellation Using Convolutional Neural Networks". *Front. Neurol*. **10**:1346. doi: 10.3389/fneur.2019.01346

13. Peng H, Gong W, Beckmann CF, Vedaldi A, Smith SM. (2021); 'Accurate brain age prediction with lightweight deep neural networks. *Med Image Anal* **68**:101871.

14. Lea Baecker, Jessica Dafflon,Pedro F. da Costa, Rafael Garcia-Dias, Sandra Vieira, Cristina Scarpazza, Vince D. Calhoun, Joao R. Sato, Andrea Mechelli, and Walter H. L. Pinaya. (2021), 'Brain age prediction: A comparison between machine learning models using region- and voxel-based morphometric data'. *Hum Brain Mapp*; **42**:2332–2346. https://doi:10.1002/hbm.25368

15. Beheshti, M. A. Ganaie, V. Paliwal, A. Rastogi, I. Razzak and M. Tanveer,(2022) "Predicting Brain Age Using Machine Learning Algorithms: A Comprehensive Evaluation," in *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. **4**, pp. 1432-1440, doi: 10.1109/JBHI.2021.3083187. https://pubmed.ncbi.nlm.nih.gov/34029201/

16. Gómez-Ramírez, J., Miguel A. Fernández-Blázquez, and Javier J. González-Rosa, (2022). "Prediction of Chronological Age in Healthy Elderly Subjects with Machine Learning from MRI Brain Segmentation and Cortical Parcellation" *Brain Sciences* **12**, no. 5: 579. https://doi.org/10.3390/brainsci12050579

17. Bruce Fischl, FreeSurfer, *Neuroimage*. (2012) Aug 15;62(**2**):774-81. doi:10.1016/j. neuroimage.2012.01.021. https://pubmed.ncbi.nlm.nih.gov/22248573/

18. Christophe Destrieux , Bruce Fischl, Anders Dale, Eric Halgren,(2010),'Automatic Parcellation of Human Cortical Gyri and Sulci using Standard Anatomical Nomenclature'. *Neuroimage* 15;53(**1**):pp1-15. https://pubmed.ncbi.nlm.nih.gov/20547229/

19. Kira, K., & Rendell, L. A. (1992b). 'A Practical Approach to Feature Selection'. Machine Learning *Proceedings 1992*. pp. 249-256. https://doi.org/10.1016/B978-1-55860-247-2.50037-1

20. Robnik-Sikonja, M., and I. Kononenko. (2003). 'Theoretical and Empirical Analysis of ReliefF and RReliefF'. *Machine Learning*, **53**, 23–69

21. Dietterich, T. G. (1997). Machine learning research: Four current directions. *AI Magazine*, 18:4, 97–136.

22. Cortes, C., & Vapnik, V. (1995). 'Support Vector |Networks'. *Machine Learning*, **20**, 273–297. https://doi.org/10.1109/64.163674.

23. Raoof Gholami, Nikoo Fakhari (2017),'Support Vector Machine: Principles, Parameters, and Applications'; *Chapter 27of Handbook of Neural Computation*., pp. 515-535. https://doi.org/10.1016/B978-0-12-811318-9.00027-2.

24. Smola, A.J., Schölkopf, B. (2004), 'A Tutorial on Support Vector Regression'. *Statistics and Computing* (**14**), 199–222 . https://doi.org/10.1023/B:STCO.0000035301.49549.88

25. Tipping, M. E. (2001). 'Sparse Bayesian learning and the relevance vector machine'. *Journal of Machine Learning Research*, 1, **211**–244. https://doi.org/10.1162/153244301527482364

26. Franke, K., & Gaser, C. (2019). '10 years of Brain-AGE as a neuroimaging biomarker of brain aging: What insights did we gain?', *Frontiers in Neurology*, 10(**789**), 1–26. https://doi.org/10.3389/FNEUR.2019.00789

27. Williams, C. K. I., & Rasmussen, C. E. (1996). 'Gaussian processes for regression'. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in neural information processing systems* (pp. 514–520). Cambridge,

28. Lizhen Lin. Niu Mu. Pokman Cheung. David Dunson.(2019) .'Extrinsic Gaussian Processes for Regression and Classification on Manifolds'. *Bayesian Anal*. 14 (**3**) 887- 906, 2019. https://doi.org/10.1214/18-BA1135

29. Ebden, M. (2008). 'Gaussian Processes for Regression: A Quick Introduction'. Retrieved from https://api.semanticscholar.org/CorpusID:124877894

30. Hui Zou and Trevor Hastie (2005), 'Regularization and Variable Selection via the Elastic Net'. J. R. Statist. Soc. B 67, Part 2, pp. 301–320.

31. Ruyi Xiao, Xinchun Cuia ,Hong Qiaob, Xiangwei Zhengc, Yiquan Zhang , Chenghui Zhanga and Xiaoli Liu.(2021) ,'Early Diagnosis Model of Alzheimer's Disease Based on Sparse Logistic Regression With the Generalized Elastic Net'. *Biomedical Signal Processing and Control*, (**66**) , 102362. https://doi.org/10.1016/j.bspc.2020.102362

32. Implementation of Elastic Net regression retrived from https://www.geeksforgeeks.org/implementation-of-elastic-net-regression-from-scratch/

33. Chris Hans, (2011)'Elastic Net Regression Modeling with the Orthant Normal Prior', *Journal of the American Statistical Association*, Volume 106, (**496**). https://doi.org/10.1198/jasa.2011.tm09241

34. Bouzid Ait-Amir, Philippe Pougnet and Abdelkhalak El Hami.(2020), '6 - Meta-Model Development', Embedded Mechatronic Systems. *Analysis of Failures, Modeling, Simulation and Optimization* , Pages 157-187, https://doi.org/10.1016/B978-1-78548-014-0.50006-2

35. Gutierrez Becker, B., Klein, T., & Wachinger, C. (2018). 'Gaussian process uncertainty in age estimation as a measure of brain abnormality'. *NeuroImage*, (**175**), 246–258. https://doi.org/10.1016/j.neuroimage.2018.03.075

36. Aycheh, H. M., Seong, J. K., Shin, J. H., Na, D. L., Kang, B., Seo, S. W., & Sohn, K.-A. (2018). 'Biological brain age prediction using cortical thickness data: A large scale cohort study'. *Frontiers in Aging Neuroscience*, 10(**252**), 1–14. https://doi.org/10.3389/fnagi.2018.00252

37. Beheshti, I., Nugent, S., Potvin, O., & Duchesne, S. (2019). 'Bias-adjustment in neuroimaging-based brain age frameworks: A robust scheme'. *NeuroImage: Clinical*, (**24**), 102063. https://doi.org/10.1016/j.nicl.2019.102063

38. De Lange, A. M. G., & Cole, J. H. (2020). 'Commentary: Correction procedures in brain-age prediction'. *NeuroImage: Clinical*, (**26**), 24–26. https://doi.org/10.1016/j.nicl.2020.102229

39. Cole, J. H., Raffel, J., Friede, T., Eshaghi, A., Brownlee, W. J., Chard, D., Ciccarelli, O. (2020). 'Longitudinal assessment of multiple sclerosis with the brain-age paradigm'. *Annals of Neurology*, (**88**), 93–105. https://doi.org/10.1002/ana.25746

40. Le, T. T., Kuplicki, R. T., McKinney, B. A., Yeh, H.-W., Thompson, W. K., Paulus, M. P., & Tulsa 1000 Investigators. (2018). 'A nonlinear simulation framework supports adjusting for age when analyzing Brain-AGE'. *Frontiers in Aging Neuroscience*, **10**(317), 1–11. https://doi.org/10.3389/fnagi.2018.00317